

REPORT DOCUMENTATION PAGE

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for review, existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Service Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE (DD-MM-YYYY) 29-04-2009		2. REPORT TYPE Final		3. DATES COVERED (From - To) 1 Oct 2005 - 31 Sep 2008	
4. TITLE AND SUBTITLE Enhancing Command Communication and Innovation with SAINT: Semantics, Adaptation, and Influence in Networked Teams				5a. CONTRACT NUMBER FA9550-06-1-0370	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
				5d. PROJECT NUMBER	
6. AUTHOR(S) David E. Goldberg Xavier Llorca				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Department of Industrial and Enterprise Systems Engineering University of Illinois at Urbana Champaign 117 Transportation Building, 104 S. Mathews Urbana, IL 61820-6242				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) USAF, AFRL Air Force Office of Scientific Research 875 N. Randolph St Arlington, VA 22203				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Penny Weaver
Uu	U	U	UU		19b. TELEPHONE NUMBER (Include area code) 217 265-7679

Final Report
**Enhancing Command Communication and Innovation with SAINT:
Semantics, Adaptation, and Influence in Networked Teams**
AFOSR Grant No. FA9550-06-1-0370

October 1, 2005 to September 31, 2008

David E. Goldberg, Xavier Llorà
IlliGAL Report No. X0029 (Limited Circulation)
February 2009

Illinois Genetic Algorithms Laboratory
Department of Industrial and Enterprise Systems Engineering
University of Illinois at Urbana-Champaign
117 Transportation Building
104 S. Mathews Avenue Urbana, IL 61801

20090520227

Final Report

**Enhancing Command Communication and Innovation
with SAINT: Semantics, Adaptation, and Influence in
Networked Teams**

AFOSR Grant No. FA9550-06-1-0370

October 1, 2005 to September 31, 2008

David E. Goldberg

Department of Industrial and Enterprise Systems Engineering
University of Illinois at Urbana Champaign
deg@illinois.edu

Xavier Llorà

National Center for Supercomputing Applications
University of Illinois at Urbana Champaign
xllora@illinois.edu

February 27, 2009

1 Objectives

The objectives of AFOSR grant no. FA9550-06-1-0370 are as follows:

1. Enhance online reflection and understanding.
2. Visualize and augment the propagation of influence.
3. Promote integrated human-computer creative collaboration.
4. Embed developed systems in larger collaborative systems and data-mining infrastructures.
5. Pilot developed systems in simulated AF environment.

Significant progress has been made on each goal. The progress is briefly reviewed in the next section.

2 Status of Effort

The key efforts conducted as part of AFOSR grant no. FA9550-06-1-0370 from October 1, 2005 to September 31, 2008 can be summarized as follows:

Fast tagging and summarization for online conversations created. A key element to enhance reflection and understanding is to reduce the amount of time required for situation awareness on online conversations. Due to the fast pace of such conversations we proposed fast methods for conversation analysis and summarization. Efficient methods for extracting relevant tags and summarizing sentences out of online conversations were reported using conversations gathered from blogs. Such methods helped reducing the awareness time required for properly grasp the direction of the conversation.

Topic tracking and transition detection helped understanding conversation dynamics. Tagging and summarization of online conversations provide a good overview of discussion. However, they do not provide an accurate blueprint of the conversation dynamic nature. Using the tagging techniques developed, we created topic tracking and transition detection techniques for online conversations. Such techniques, when properly coupled with the right visualization tools, allow us to link relevant topics and time. They also allow us identifying when online conversations move from one topic to another, painting an accurate blueprint of the analyzed online conversation.

Pilot studies on identifying semantic building blocks on online conversations. Online communication takes advantage of the implicit freedom that the communication channel provides. People tend to create and use new terms and expressions, giving concise meaning to groupings of such newly created language artifacts. Enhancing the semantic understanding of online conversations requires the ability to identify these newly created constructs. Following the ideas and techniques that lead to the creation of competent genetic algorithms (Goldberg, 2002), we have proposed an initial model to identify semantic building blocks on online conversations. Pilot studies showed that for reasonable size conversations, blocks of semantically equivalent concepts were properly identified.

Visualization for topic overlap successfully identifies groups of expertise. Enhancing the visualization to boost reflection on online conversations greatly benefited from the advances developed on tagging, topic tracking and transition detection. However, any given conversation involves people. Thus, SAINT also focused mapping topics to people. Such a mapping allows us to better characterizing the interests and expertise of participants in any given online conversation. With this goal in mind, we develop complementary analysis and visualization tools to SAINTs influence diffusion models able to map and visualize key topics to people. Given any conversation we are able to present and visualize the degree of overlap between the participants. Such analysis and visualization tools successfully identified significant groups of interest and expertise on real-world online focus groups conducted with commercial partners.

Evaluation relaxation for interactive genetic algorithms invented. Interactive genetic algorithms are able to building models of user interaction when embedded on collaborative environments. One of the bottlenecks of interactive GAs where users evaluate the relative quality of candidate solutions is user fatigue. Building models of human-collaboration requires mitigating the effects of repetitive interactions. A principled method for eliminating user fatigue was invented and significant speedups were reported. In essence, the method relies on a surrogate induced from a partial ordered graph of user preferences. The surrogate is used for optimization to partially replace expensive user evaluations.

Standardized conversation format allows homogeneous access for analytic components. Online conversation may take different forms. Email, instant messaging, blogs, and wikis me-

dia provide their own standardized representation of the conversations. Such heterogeneous representation makes it harder to allow conversation analytics across the set of present and future online communication media. Thus, we defined a basic ontology to describe online conversations independently of the originating media. Based on RDFs, SAINTs conversation standardization allows unified access for the analytic tools to online conversations, regardless of the originating media. Standardized online conversations are stored on efficient RDF repositories for proper storage and management.

SAINTs non-intrusive pluggable technology pilot developed and deployed. A key goal of SAINT technology is its pluggable and non-intrusive nature. Connecting to existing legacy systems or publicly available online conversation source should not change or interfere with the current IT organizational setup. Likewise, standardized access and interoperability with SAINTs components are also part of such an endeavor. We have released a first pilot of SAINTs architecture able to interoperate with common technologies and products facilitating the analysis of RSS feeds, blogs, email, and discussion boards. Also, such pilot allows third party application to have access to its analytic components via web services. SAINT architecture successfully powered conversation analysis on different scenarios involving academic and commercial partners.

Better understanding of online creative sessions. Advances on the integration of: (1) fast tagging and summarization of its contents, (2) topic tracking and topic transition detection, (3) measures of participants impact and influence, and (4) techniques for identifying semantic building blocks. Building on the top of these elements we focused on quantifying and advancing the principled understanding of the performance boost introduced by modeling online conversations as evolutionary computation processes. Our results showed that modeling the conversation as a selectorecombinative process could increase the activity of participants (up to 2.6 times more messages). Also, modeling an online conversation after human-based genetic algorithms led to richer discussions with more distinct ideas (3 times more) without penalizing quality. Moreover, to achieve the same number and quality and number of ideas using a selectorecombinative required less time—approximately one third of the time required by traditional means—hence reducing participants fatigues.

Understanding human-computer interaction modeling. Active interactive genetic algorithms (aiGAs) showed that building accurate models of human-computer interaction was possible. aiGAs also revealed themselves as powerful tools to combat user fatigue by actively modeling user preferences. aiGAs could provide speedups between 3 and 7 times. Such speedups were possible by reducing generating educated guesses after user preferences thanks to aiGA interaction modeling. Despite these impressive results, we still lacked a principled understanding of how to size the human-computer interaction. That is, how many distinct solutions and comparisons should we present to a user. To answer this question our research involved exploring graph-theoretic measures to understand the key element of the sizing process: The partial-ordering graph. Using measures such as the density of a graph, we identified the existence of a lower bound in the density of the partial-ordering graph to achieve and accurate interaction modeling (power law based). The density boundary could then be used to correctly predict when a particular interaction sizing will fail or succeed. It also helped on conducting informed design decision to improve the modeling of aiGAs, and explain why some models could not work.

Fitness inheritance boosting genetics-based machine learning. A key element to the effi-

ciency of SAINT modeling techniques is the ability to learn accurate models quickly. To guarantee such efficiency we brought competent genetic algorithms lessons into the genetics-based machine learning (GBML) realm—a learning technique commonly use in the SAINT project. SAINT efforts focused on the reduction of the rule-matching time required by GBML techniques which may take up to 99% of the overall execution time as we have shown. Building on efficient enhancement techniques for competent genetic algorithms, we adapted fitness inheritance technique to allow building surrogate functions that compute the accuracy of the rule matching process—which often requires the ability to deal with overlapping building blocks. Research results showed that computing the matching accuracy using surrogate functions of the matching process could be obtained when overlapping building blocks could be expressed. Once such surrogate functions was created, there was no need to explicitly execute the rule matching mechanics for a given problem, since the accuracy could be estimated using the obtained surrogate function.

Meandre: Flexibility via data-intensive flow. After the creation of a non-intrusive pluggable pilot architecture for SAINT—as reported previously—we started developing the next generation of flexible infrastructure required for the SAINT vision. Building on semantic-web technologies, Meandre is a semantic-driven data-intensive flow execution architecture. Meandre was originally conceived as a flexible, reconfigurable, and reusable infrastructure for data-intensive applications. Meandre can scale from a single laptop to large clusters transparently. It allows the creation of components using several languages (basic computation units), which can later be orchestrated together in a flow (a complex computation task). The overall design heavily relies on semantic-web technologies and service oriented architectures. Meandre has also given birth to the ZigZag programming language for data-intensive flow programming. The SAINT team has also migrated all the research algorithms for conversation analysis and visualization to Meandre components. Doing so, SAINT fulfills one of the key elements of its original proposal, to minimize the gap for transferring successful research onto reusable products. For more information please see <http://seasr.org/meandre>.

Meandre beyond SAINT. Meandre infrastructure, originally designed and developed under SAINT, has been adopted by the National Center for Supercomputing Applications as its main infrastructure for conducting data and text mining for the humanities under the SEASR project. The SEASR project is developing Meandre data and text mining components, extending the ecology of components that can be quickly reused to face new analytic tasks. Other humanities projects like Monk and Nema have also recently adopted Meandre. The evolutionary computation package (E2K) widely used in SAINT has also been migrated to the new Meandre infrastructure.

Semantic description of online conversations and visualizations. Online conversation may take different forms. Email, instant messaging, blogs, and wikis media provide their own standardized representation of the conversations. During the last reporting period, SAINT focused on creating a homogeneous description format, generic enough to engulf all these heterogeneous sources. SAINT has continued this effort during the reported period summarized by this report by consolidating these descriptions under the RDF umbrella. Ontological descriptions of conversations, and also now analysis results, and visualization formats are available under RDFS or OWL. Such standardization greatly simplified the curation, storage, and exchange of such valuable information.

Rapid SAINT distribution and deployment via hardware virtualization in the cloud.

Rapid deployment of SAINT technology requires the ability to run on available (or assigned) resources and minimize the configuration and installation efforts. SAINT analytic and visualization solutions should be deployable quickly and reliably across multiple OS including cloud-computing scenarios. To achieve such a goal, SAINT relies on hardware virtualization as the key to maximize deployment portability. We created several virtual appliances containing: (1) SAINTs online discussion community software, (2) the analytics infrastructure, (3) specialized pluggable conversation extractors, and (4) publishing capabilities for analytic and visualization services. Each specialized virtual appliance contains a canned version of SAINT ready-to-use technology making tedious installation and configuration processes unnecessary. Virtual appliances streamline the deployment of SAINT by providing a ready and scalable off-the-shelf solution.

3 Personnel Supported

This section details the individuals supported on this project.

Faculty supported. Professor David E. Goldberg, the principal investigator, was supported during the summers of 2007. Doctor Xavier Llorà, CO-PI, was supported for some project-related traveling during 2007.

Other affiliated visitors, postdoctoral personnel. The following is a list of visiting faculty or postdocs affiliated with the project. Unsupported affiliates may have had some travel or incidental expenses paid by the project:

1. Professor Pier Luca Lanzi (Politecnico di Milano, Italy)
2. Kazuisha Inaba (University of Tokyo, Japan)
3. Dr. Noriko Imafuji Yasui (University of Illinois at Urbana-Champaign)
4. Professor Avi Ostfeld (Technion, Israel Institute of Technology)
5. Albert-Orriols Puig (Ramon Llull University, Barcelona, Spain)
6. Luis de la Ossa (University of Castilla la Mancha, Spain)
7. Hiroshi Tamura (Hakuhodo Corporation and University of Tokyo, Japan)
8. Yuichi Washida (Hakuhodo Corporation and University of Tokyo, Japan)

Graduate student affiliates. The following is a list of graduate students supported or affiliated with the project (source of funding in parentheses). Unsupported affiliates may have had some travel or incidental expenses paid by the project:

1. Thyago Sellmann Pinto Cesar Duque (AFOSR)
2. Osvaldo Gomez (Fullbright)
3. Kazuisha Inaba (AFOSR)
4. Kumara Sastry (MatSE)
5. Paul Winward (AFOSR)
6. Tian-Li Yu (AFOSR)

Undergraduate student affiliates. The following is a list of graduate students supported or affiliated with the project. Unsupported affiliates may have had some travel or incidental expenses paid by the project:

1. Takaoki Ueda (AFOSR)
2. Dae Hoon Park (AFOSR)
3. Jordan Hawker (AFOSR)

4 Publications for 2005

Published

- Butz, M. V., D. E. Goldberg, and P. L. Lanzi, Computational Complexity of the XCS Classifier System. In L. Bull and T. Kovacs (Eds), *Foundations of Learning Classifier Systems*, 91–126 (2005).
- Butz, M. V., D. E. Goldberg, and P.-L. Lanzi, Gradient Descent Methods in Learning Classifier Systems: Improving XCS Performance in Multistep Problems. *IEEE Transaction on Evolutionary Computation*, 9(5), 452–473 (2005).
- Butz, M. V., M. Pelikan, X. Llorà, and D. E. Goldberg, Extracted global structure makes local building block processing effective in XCS. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2005)*, 655–662 (2005). [Best paper award nominee, Estimation of Distribution Algorithm track].
- Butz, M. V., K. Sastry, and D. E. Goldberg. Strong, Stable, and Reliable Fitness Pressure in XCS due to Tournament Selection. *Genetic Programming and Evolvable Machines*, 6(1), 53–77 (2005).
- Goldberg, D. E., John H. Holland, Facetwise Models, and Economy of Thought. In L. Booker, S. Forrest, M. Mitchell, and R. Riolo (Eds.), *Perspectives on Adaptation in Natural and Artificial Systems* (pp. 57–69). New York: Oxford University Press (2005).
- Goldberg, D. E. Foreword. In Abe, A., and Y. Ohsawa (Eds.), *Readings in Chance Discovery* (pp. i–ii). Adelaide, Australia: Advanced Knowledge International (2005).
- Lanzi, P.-L., D. Loiacono, S. W. Wilson, and D. E. Goldberg, XCS with Computable Prediction for the Learning of Boolean Functions. *Proceedings of the 2005 Congress on Evolutionary Computation (CEC2005)*, 588–595 (2005).
- Lanzi, P.-L., D. Loiacono, S. W. Wilson, and D. E. Goldberg, XCS with Computable Prediction in Continuous Multistep Environments. *Proceedings of the 2005 Congress on Evolutionary Computation (CEC2005)*, 2032–2039 (2005).
- Lanzi, P.-L., D. Loiacono, S. W. Wilson, and D. E. Goldberg, Extending XCSF Beyond Linear Approximation. *Proceedings of the Genetic and Evolutionary Computation Conference*, 1827–1834 (2005).
- Lanzi, P.-L., D. Loiacono, S. W. Wilson, and D. E. Goldberg, XCS with Computable Prediction in Multistep Environments. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2005)*, 1859–1866 (2005).
- Llorà, X., K. Sastry, and D. E. Goldberg, The compact classifier system: Motivation, analysis and first results. *Proceedings of the 2005 Congress on Evolutionary Computation*, 1, 596–603, (2005).
- Llorà, X., K. Sastry, D. E. Goldberg, The Compact Classifier System: Motivation, Analysis, and First Results. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2005)* [poster], 1993–1994 (2005).

- Llorà, X., K. Sastry, D. E. Goldberg, A. Gupta, and L. Lakshmi, Combating User Fatigue in iGAs: Partial Ordering, Support Vector Machines, and Synthetic Fitness. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2005)*, 1363–1370 (2005).
- Matsumura, N., D. E. Goldberg, and X. Llorà, Mining Social Networks in Message Boards. *Proceedings of the Symposium on Conversational Informatics for Supporting Social Intelligence*, 18–27 (2005).
- Pelikan, M., Sastry, K. Goldberg, D. E. (2005). Multiobjective hBOA, clustering, and scalability. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2005)*, 663–670 (2005).
- Qurin, A., J. Koreczak, M. V. Butz, and D. E. Goldberg, Analysis and Evaluation of Learning Classifier Systems Applied to Hyperspectral Image Classification. *Proceedings of the Fifth International Conference on Intelligent Systems Design and Applications*, 280–285 (2005).
- Sastry, K., M. Pelikan, D. E. Goldberg, Limits of scalability of multiobjective estimation of distribution algorithms. *Proceedings of the 2005 Congress on Evolutionary Computation*, 3, 2217–2224, (2005).
- Sastry, K., H. A. Abbass, D. E. Goldberg, and D. D. Johnson, Sub-structural Niching in Estimation of Distribution Algorithms. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2005)*, 671–678 (2005). [Best paper award nominee, Estimation of Distribution Algorithms track].
- Sastry, K., D. D. Johnson, D. E. Goldberg, and P. Bellon, Genetic Programming for Multi-Timescale Modeling: A Proof-of-Principle Study. *Physics Review B*, 72(8), 085438, (2005).
- Yassine, A., D. E. Goldberg, and T.-L. Yu, Simple Models of Hierarchical Organizations. *Proceedings of the Academy of Management*, (2005).
- Yu, T.-L., K. Sastry, and D. E. Goldberg, Online population size adjusting using noise and sub-structural measurements. *Proceedings of the 2005 Congress on Evolutionary Computation Conference*, 3, 2491–2498, (2005).
- Yu, T.-L., K. Sastry, and D. E. Goldberg, Linkage Learning, Overlapping Building Blocks, and a Systematic Strategy for Scalable Recombination. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2005)*, 1217–1224 (2005).

5 Publications for 2006

Published

- Alías, F. , X. Llorà, L. Formiga, K. Sastry, and D. E. Goldberg, Efficient interactive weight tuning for TTS synthesis: Reducing user fatigue by improving user consistency. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2006)*, 1, 865–868, (2006).
- Butz, M.V., M. Pelikan, X. Llorà, and D. E. Goldberg, Automated Global Structure Extraction For Effective Local Building Block Processing in XCS. *Evolutionary Computation*. 14(3), 345–380 (2006).
- Goldberg, D. E., *The Entrepreneurial Engineer*. New York, NY: Wiley (2006).
- Harik, G. R., F. G. Lobo, and K. Sastry, Linkage Learning via Probabilistic Modeling in the ECGA. In Pelikan, M., Sastry, K., Cantú-Paz, E. (Eds.), *Scalable Optimization via Probabilistic Modeling: From Algorithms to Applications* (pp 39–61). Berlin: Springer (2006).

- Lanzi, P. L., D. Loiacono, S. W. Wilson, and D. E. Goldberg, Classifier Prediction Based on Tile Coding. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2006)*, 1497–1504 (2006). [Best paper award, Learning Classifier Systems and other Genetics-Based Machine Learning track].
- Lanzi, P. L., D. Loiacono, S. W. Wilson, and D. E. Goldberg, Prediction Update Algorithms for XCSF: RLS, Kalman Filter, and Gain Adaptation. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2006)*, 1505–1512 (2006).
- Lima, C.F., M. Pelikan, K. Sastry, M. Butz, D. E. Goldberg, and F. G. Lobo, Substructural Neighborhoods for Local Search in the Bayesian Optimization Algorithm. *Proceedings of the Parallel Problem Solving from Nature (PPSN IX)*, 232–241 (2006).
- Llorà, X., Evolution to Knowledge, *ACM SIGEvolution Newsletter*, Volume 1 , Issue 3, pp. 10-17, (2006).
- Llorà, X., D. E. Goldberg, Y. Ohsawa, N. Matsumura, Y. Washida, H. Tamura, M. Yoshikawa, M. Welge, L. Auvil, D. Searshmith, K. Ohnishi, and C.-J. Chao, Innovation and Creativity Support via Chance Discovery, Genetic Algorithms, and Data Mining. *New Mathematics and Natural Computation*, 2(1), 85–100 (2006).
- Llorà, X., and K. Sastry, Fast rule matching for learning classifier systems via vector instructions. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2006)*, 1513–1520, (2006).
- Llorà, X., K. Sastry, F. Alías, D. E. Goldberg, and M. Welge, Analyzing active interactive genetic algorithms using visual analytics. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2006)* [poster], 1417–1418, (2006).
- Matsumura, N., Goldberg, D.E., and Llorà, X. (2006). Communication gap management for fertile community, *Journal of Soft Computing*, Volume 11 , Issue 8, pp. 791798, ACM press.
- Pelikan, M. and D. E. Goldberg (2006). Hierarchical Bayesian Optimization Algorithm. In M. Pelikan, K. Sastry, and E. Cantú-Paz (Eds), *Scalable Optimization via Probabilistic Modeling: From Algorithms to Applications*. Berlin: Springer (2006).
- Pelikan, M., A. K. Hartmann, and K. Sastry, Hierarchical BOA, Cluster Exact Approximation, and Ising Spin Glasses. *Parallel Problem Solving from Nature (PPSN IX)*, 122–131 (2006).
- Pelikan, M., K. Sastry, M. V. Butz, D. E. Goldberg, Hierarchical BOA on random decomposable problems. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2006)* [poster], 431–432, (2006).
- Pelikan, M., K. Sastry, M. V. Butz, and D. E. Goldberg, Performance of Evolutionary Algorithms on Random Decomposable Problems. *Parallel Problem Solving from Nature (PPSN IX)*, 788–797 (2006).
- Pelikan, M., K. Sastry, E. Cantú-Paz, (Eds). *Scalable Optimization via Probabilistic Modeling: From Algorithms to Applications*, Berlin: Springer (2006).
- Pelikan, M., K. Sastry, and D. E. Goldberg. Multiobjective Estimation of Distribution Algorithms. In M. Pelikan, K. Sastry, and E. Cantú-Paz (Eds), *Scalable Optimization via Probabilistic Modeling: From Algorithms to Applications* (pp 223–248). Berlin: Springer (2006).
- Pelikan, M., K. Sastry, and D. E. Goldberg, Sporadic Model Building for Efficiency Enhancement of Hierarchical BOA. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2006)*, 405–412 (2006).

- Santarelli S., T.-L. Yu, D. E. Goldberg, E. Altshuler E., T. O'Donnell, H. Southall, and R. Mailoux, Military Antenna Design Using Simple and Competent Genetic Algorithms. *Mathematical and Computer Modeling*, 43, 990–1022 (2006).
- Sastry, K., D.D. Johnson, A. L. Thompson, D. E. Goldberg, T. J. Martinez, J. Leiding, and J. Owens, Multiobjective genetic algorithms for multiscaling excited state direct dynamics in photochemistry. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2006)*. 1745–1752, (2006). [Best paper award in Real World Applications track] [Silver “Humies” award at the Human Competitive Results Competition].
- Sastry, K., D. E. Goldberg, G. Kendall, Genetic algorithms: A tutorial. In E. Burke, and G. Kendall (Eds), *Introductory Tutorials in Optimization, Search and Decision Support Methodologies*. Berlin: Springer (2006).
- Sastry, K., C. F. Lima, and D. E. Goldberg, Evaluation relaxation using substructural information and linear estimation. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2006)*. 419–426, (2006).
- Sastry, K., M. Pelikan, and D. E. Goldberg (2006). Efficiency Enhancement of Estimation of Distribution Algorithms. In M. Pelikan, K. Sastry, and E. Cantú-Paz (Eds), *Scalable Optimization via Probabilistic Modeling: From Algorithms to Applications* (pp 161–185). Berlin: Springer (2006).
- Sastry, K., P. Winward, D. E. Goldberg, and C. F. Lima, Fluctuating crosstalk as a source of deterministic noise and its effects on GA scalability. *Applications of Evolutionary Computing EvoWorkshops2006: EvoBIO, EvoCOMNET, EvoHOT, EvoIASP, EvoInteraction, EvoMUSART, EvoSTOCK*, 740–751, (2006).
- Yu, T.-L., and D. E. Goldberg, Conquering Hierarchical Difficulty by Explicit Chunking: Chromosome Compression. *Proceedings of the 2006 Genetic and Evolutionary Computation Conference*, 1385–1392 (2006).
- Yu, T.-L., S. Santarelli, and D. E. Goldberg, Military Antenna Design Using Simple Genetic Algorithm and hBOA. In M. Pelikan, K. Sastry, and E. Cantú-Paz (Eds), *Scalable Optimization via Probabilistic Modeling: From Algorithms to Applications*. Berlin: Springer (2006).
- Yu, T.-L., A. Yassine, and D. E. Goldberg, Double Duty: Genetic Algorithms for Organizational Design and Genetic Algorithms Inspired by Organizational Theory. In J.-P. Rennard (Ed.), *Handbook of Nature Inspired Computing for Economy and Management*, (Chapter 28) Hershey, PA: Idea Group Reference (2006).

6 Publications for 2007

- Bacardit, J., D. E. Goldberg, and M. V. Butz, Improving the Performance of a Pittsburgh Learning Classifier System Using a Default Rule. In XCS. In T. Kovacs, X. Llorà, K. Takadama, P. L. Lanzi, W. Stolzmann, and S. W. Wilson (Eds.) *Learning Classifier Systems*, 299–318 (2007).
- Bacardit, J., M. Stout, K. Sastry, X. Llorà, N. Krasnogor, and J. D. Hirst, Automated alphabet reduction method with evolutionary algorithms for protein structure prediction. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2007)*, 346–353 (2007).
- Butz, M. V., D. E. Goldberg, and P. L. Lanzi, Effect of Pure Error-Based Fitness. In XCS. In T. Kovacs, X. Llorà, K. Takadama, P. L. Lanzi, W. Stolzmann, and S. W. Wilson (Eds.) *Learning Classifier Systems*, 105–115 (2007).

- Butz, M. V., D. E. Goldberg, P. L. Lanzi, and K. Sastry, Problem Solution Sustenance in XCS: Markov Chain Analysis of Niche Support Distributions and Consequent Computational Complexity. *Genetic Programming and Evolvable Machines*, 8, 5–37. (2007).
- Chen, Y.-p., T.-L. Yu, K. Sastry, and D. E. Goldberg, A Survey of Linkage Learning Techniques in Genetic and Evolutionary Algorithms. IlliGAL Technical Report No. 2007014, University of Illinois at Urbana-Champaign (2007).
- Goldberg, D. E., K. Sastry, and X. Llorà, Toward routine billion-variable optimization using genetic algorithms. *Complexity*, 12(3), 27–29 (2007).
- Hauschild, M., M. Pelikan, C. F. Lima, and K. Sastry, Analyzing Probabilistic Models in Hierarchical BOA on Traps and Spin Glasses. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2007)*, 523–530 (2007).
- Lanzi, P.-L., D. Loiacono, S. W. Wilson, and D. E. Goldberg, Generalization in the XCSF Classifier System: Analysis, Improvement, and Extension. *IEEE Transactions on Evolutionary Computation*, 15:2, 133–168 (2007).
- Llorà, X., Priya, A., and Bhargava, R., Observer-Invariant Histopathology using Genetics-Based Machine Learning. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2007)*, 2098–2105 (2007).
- Llorà, X., K. Sastry, and D. E. Goldberg, Binary Rule Encoding Schemes: A Study Using the Compact Classifier System. In XCS. In T. Kovacs, X. Llorà, K. Takadama, P. L. Lanzi, W. Stolzmann, and S. W. Wilson (Eds.) *Learning Classifier Systems*, 41–60 (2007).
- Llorà, X., K. Sastry, T.-L. Yu, and D. E. Goldberg, Do not match, Inherit: Fitness surrogates for genetics-based machine learning techniques. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2007)*, 1798–1805 (2007).
- Llorà, X., N. I. Yasui, M. Welge and D. E. Goldberg, Human-Centered Analysis and Visualization Tools for the Blogosphere. *Proceedings of the 2007 Digital Humanities Conference*, 127–130 (2007).
- Llorà, X., N. I. Yasui, and D. E. Goldberg, Analyzing trends in the blogosphere using human-centered analysis and visualization tools. *Proceedings of the International Conference on Weblogs and Social Mining (ICWSM 2007)* 285–286 (2007).
- Matsumura, N., D. E. Goldberg, and X. Llorà, Communication Gap Management for Fertile Community. *Journal of Soft Computing*, 11(8), 791–798 (2007).
- Minqiang, L., D. E. Goldberg, K. Sastry, and T.-L. Yu, Real-coded ECGA for solving decomposable real-valued optimization problems. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2007)* [poster], (2007).
- Minqiang, L., D. E. Goldberg, K. Sastry, and T.-L. Yu, Performance analyses of factorization based on Gaussian PDF in RECGA. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2007)* [poster], (2007).
- Orriols-Puig, A., D. E. Goldberg, K. Sastry, and E. Bernadó-Mansilla, Modeling XCS in class imbalances: Population sizing and parameter settings. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2007)*, 1838–1845 (2007).
- Orriols-Puig, A., K. Sastry, P. L. Lanzi, D. E. Goldberg, and E. Bernadó-Mansilla, Modeling selection pressure in XCS for proportionate and tournament selection. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2007)*, 1846–1853, (2007).

- Pelikan, M., K. Sastry, and D. E. Goldberg, Sporadic Model Building for Efficiency Enhancement of Hierarchical BOA. *Genetic Programming and Evolvable Machines*, 9(1). 53–84. (2007).
- Ueda, T., N. I. Yasui, X. Llorà, K. Sastry, and D. E. Goldberg, Discovering Building Blocks for Human Based Genetic Algorithms. *Artificial Neural Networks In Engineering*, (2007).
- Sastry, K., and D. E. Goldberg, Let's get ready to rumble redux: Crossover versus mutation head to head on exponentially scaled problems. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2007)*, 1380–1387 (2007).
- Sastry, K., D. E. Goldberg, and X. Llorà, Towards billion bit optimization via parallel estimation of distribution algorithm. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2007)*, 577–584 (2007). [Best paper award, Estimation of Distribution Algorithms track].
- Sastry, K., D. D. Johnson, and D. E. Goldberg, Scalability of a hybrid extended compact genetic algorithm for ground state optimization of clusters. *Materials and Manufacturing Processes*, 22(5), 570–576. (2007).
- Sastry, K., D. D. Johnson, A. L. Thompson, D. E. Goldberg, T. J. Martinez, T. J., J. Leiding, and J. Owens, Optimization of Semiempirical Quantum Chemistry Methods via Multiobjective Genetic Algorithms: Accurate Photochemistry for Larger Molecules and Longer Time Scales *Materials and Manufacturing Processes*, 22(5), 553–561 (2007).
- Sastry, K., M. Pelikan, and D. E. Goldberg, Analysis of Ideal Recombination on Random Decomposable Problems. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2007)*, 1388–1395 (2007). [Best paper award nominee, Genetic Algorithms track].
- Yasui, N. I., X. Llorà, D. E. Goldberg, Y. Washida, and H. Tamura, Delineating Topic and Discusant Transitions in Online Collaborative Environments. *Proceedings of the International Conference on Enterprise Information Systems (ICEIS)*, (2007).
- Yu, T.-L., K. Sastry, D. E. Goldberg, and M. Pelikan, Population sizing for entropy-based model building in genetic algorithms. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2007)*, 601–608 (2007). [Best paper award nominee, Estimation of Distribution Algorithms track].
- Yu, T.-L., K. Sastry, and D. E. Goldberg, Population sizing to go: Online adaptation using noise and substructural measurement. In F. G. Lobo, C. F. Lima, and Z. Michalewicz (Eds.), *Parameter Settings in Evolutionary Algorithms*. Berlin: Springer (2007).
- Yu, T.-L., A. Yassine, and D. E. Goldberg. An Information Theoretic Method for Developing Modular Architectures Using Genetic Algorithms. *Research in Design Engineering*, 91–109 (2007).

7 Publications for 2008

- Butz, M.V., P.L. Lanzi, X. Llorà, D. Loiacono, An Analysis of Matching in Learning Classifier Systems, *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2008)*, 1349–1356 (2008).
- Duque, T., D.E. Goldberg, K. Sastry, Improving the efficiency of the extended compact genetic algorithm. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2008)*, 467–468 (2008).

- Lanzi, P.L., L. Nichetti, K. Sastry, D. Voltini, D. E. Goldberg, Real-Coded Extended Compact Genetic Algorithm based on Mixtures of Models *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2008)*, 335–358 (2008).
- Llorà, X., B. Acs, L. S. Auvil, B. Capitanu, M. E. Welge, D. E. Goldberg, Meandrc: Semantic-Driven Data-Intensive Flows in the Clouds, Fourth IEEE International Conference on eScience, 238–245 (2008)
- Llorà, X., N.I. Yasui, D. E. Goldberg, Graph-Theoretic Measure for Active iGAs: Interaction Sizing and Parallel Evaluation Ensemble *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2008)*, 985–992 (2008).
- Sato Y., D.E. Goldberg, K. Sastry, Improving Small Population Performance under Noise with Viral Infection + Tropism, *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2008)*, 1143–1144 (2008).
- Saruwatari, S., X. Llorà, N.I. Yasui, H. Tamura, K. Sastry, D.E. Goldberg, Speeding Online Synthesis via enforced selecto-recombination. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2008)*, 1635–1642 (2008). [Best paper award nominee, Real World Applications].
- Yasui, N. I., X. Llorà, S. Saruwatari, and D. E. Goldberg. Facilitation Support for On-line Focus Group Discussions by Message Feature Map. *Proceedings of the International on Enterprise Information Systems* (2008).

8 Publications for 2009

In Press

- Llorà, X., and D. E. Goldberg, The Innovation Pump: Supporting Creative Processes in Collaborative Engineering. *International Journal on Collaborative Engineering* (2009)

Under Revision

- Duque, T. S. P. C. , and D. E. Goldberg, ClusterMI: Building Probabilistic Models using Hierarchical Clustering and Mutual Information, Submitted to *Genetic and Evolutionary Computation Conference (GECCO)*, 2009.
- Goldberg, D. E., A. Yassine, and T.-L. Yu, Decision, Execution and Team Size: Insights from Simple Quantitative Models. *IEEE Transactions on Management* (2009).
- Llorà, X. Data-Intensive Computing for Competent Genetic Algorithms: A Pilot Study using Meandre. Submitted to *Genetic and Evolutionary Computation Conference (GECCO)*, 2009.
- Moreno-Torres, J. G., X. Llor, and D.E Goldberg, Binary Representation in Gene Expression Programming: Towards a Better Scalability Submitted to *Genetic and Evolutionary Computation Conference (GECCO)*, 2009.
- Pelikan, M., K. Sastry, M. V. Butz, and D. E. Goldberg, Genetic and evolutionary algorithms on random additively decomposable problems. *Evolutionary Computation Journal*, (2009).
- Sastry, K., M. Pelikan, and D. E. Goldberg, Efficiency Enhancement of Genetic Algorithms by Building an Internal Probabilistic Model of Fitness. *Evolutionary Computation Journal* (2009).

Vallim, R. M. M., T. S. P. C. Duque, D. E. Goldberg, and A. C. P. L. F. Carvalho, The Multi-label OCS with a Genetic Algorithm for Rule Discovery: Implementation and First Results, Submitted to *Genetic and Evolutionary Computation Conference (GECCO)*, 2009.

Published

Llorà, X., K. Sastry, C. F. Lima, F. G. Lobo, D. E. Goldberg, Linkage Learning, Rule Representation, and the χ -ary Extended Compact Classifier System Compact Classifier System, In *Bacardit J., Bernadó, Butz, M, T. Kovacs, X. Llorà, K. Takadama*, (Eds.) *Learning Classifier Systems*, (2009)

9 Interactions and Transitions

This section lists meeting participation, presentations, and transitions.

Meeting Participation and Presentation

All conference papers above represent presentations by Professor Goldberg, Doctor Xavier Llorà, their affiliates, or Professor Goldberg's and Doctor Llorà's students. Professor Goldberg gave the following keynote speeches:

From Optimization to Invention Machine: The Genetic Path to Competent & Efficient Computation IEEE International Conference on Image Processing, Genoa, Italy (2005).

Little Models, Big Results. 18th Joint Australian Conference on Artificial Intelligence, and the Australian Conference on Artificial Life, Sydney, Australia (2005).

Solving Larger, Faster, and Harder: The Incredible Story of Supermultiplicative Speedups. Frontiers of Computational Science, Nagoya, Japan, 2005.

Reflections of a Keen Modeler. Workshop on Optimization by Building and Using Probabilistic Models, San Louis, Missouri, 2007.

The Innovation Imperative, University of West England (2007).

What engineers don't learn and how they don't learn it, Society for Philosophy and Technology, American Philosophical Association, 2008.

Professor Goldberg has also founded and helped organizing the following activities.

Engineering and Technology Studies at Illinois. Professor Goldberg together with professor Michael Loui started organizing in 2007 a seminar series focussed on studies revolving engineering and technology. Although everyone benefits from the products of engineering creativity and productivity, engineering and technology are seldom studied by scholars as human activities. Thus engineers, engineering, and technology are underrepresented on the intellectual landscape in a way that is inconsistent with their importance in the twenty-first century. The study of engineering and technology from philosophical, historical, cultural, behavioral, sociological, and engineering perspectives is essential to address important contemporary problems at the interface of technology and society. Cutting-edge research have been invited to share their vision in the series. Two editions of the series has been conducted in 2007 and 2008. More information can be found at <http://www.illigal.uiuc.edu/web/etsi/>.

Workshop on Philosophy & Engineering. The workshops mission is to encourage reflection on engineering, engineers, and technology by philosophers and engineers alike. A meeting was held at the Engineering Systems Division of MIT on 19-20 October 2006 by the Philosophy and Engineering Planning group headed by Taft Broome, Howard University. At that meeting, a decision was reached to hold a Workshop on Philosophy and Engineering in 2007. WPE-2007 was held at the Technical University of Delft, 29-31 October 2007 (Monday-Wednesday) with 81 participants, 40 paper presentations, 5 posters, 6 invited speakers, and attendees from 14 different countries. The second edition (2008) was conducted at The Royal Academy of Engineering, London, England. More information can be found at <http://www.illegal.uiuc.edu/web/wpe/>.

Additionally, Doctor Llorà gave the following keynote speeches, invited talks, and panels:

The Evolutionary Path to Innovation and Creativity [keynote], Congreso Mexicano de Computación Evolutiva (CONCEV), Aguas Calientes, Mexico, May, 2005.

Evolutionary Tools for Human-Innovation and Creativity [panel speech], DTM-2006 Synthesis and Innovation, Philadelphia, September, 2006.

Information Technologies for Supporting Human Innovation and Creativity [keynote], Institute of People-Centered Computation (IPCC), Bristol, UK, December, 2006.

Innovation and Creativity Support for Real-World Online Marketing [invited talk], Boeing Corporation, Seattle, March, 2007.

Evolution and Information Technologies to Support Innovation and Creativity [invited talk], Cubes for Innovation: Information Systems for Design and Marketing, Tokyo, Japan, March, 2007.

Meandre: Semantic-Web-Driven Data Flow Infrastructure for SEASR [invited talk], Service Oriented Computing in the Humanities, London, UK, 2007.

Facilitation of Online Discussion Groups, Open University of Catalonia, Barcelona, Spain, 2008.

Transitions

Meandre open source license approved. The Meandre infrastructure, an semantic-driven data-intensive flow architecture has been approved to be released under University of Illinois/National Center for Supercomputing Applications OS license. Meandre is currently powering all the analysis tools used in SAINT.

Meandre adopted for center efforts in humanities. Meandre infrastructure, originally designed and developed under SAINT, has been adopted by the National Center for Supercomputing Applications as its main infrastructure for conducting data and text mining for the humanities under the SEASR project. More information can be obtained at <http://seasr.org> and <http://seasr.org/meandre>

DISCUS/SAINT license for pilot extended. The current license to use DISCUS/SAINT technology by Hakuhodo Inc. (the second largest advertisement company in Japan) has been extended, allowing Hakuhodo to continue with its commercial trials of DISCUS/SAINT technology. Also, a open source usage license has been approved to allow people to access DISCUS/SAINT demos and basic technology. The licensing is currently under revision to allow the inclusion of Fork Inc. (a Japanese software development firm).

E2K open source license approved. Evolution to knowledge, an evolutionary-driven package for optimization and machine learning has been released under OS license for academic usage. It provides basic GA related algorithms used in SAINT (e.g. eCGA and DSM technology).

Nextumi releases ShareThis. With venture capital provided by IllinoisVentures and Blue Chip Venture Capital, Nextumi Inc., a web 2.0 startup company that simplifies sharing of photos, videos, urls and other content between people across different devices released its first product ShareThis. Nextumi licensed AFOSR sponsored DISCUS/SAINT (patents) to help to create consumer-adaptive technology.

Entrepreneurial engineer published. Professor Goldberg published a new book, *The Entrepreneurial Engineer* (Wiley, 2006) for engineering students and practicing engineers to help develop those personal, interpersonal, and organizational skills necessary today.

Advances at the frontiers of Learning Classifier Systems. Doctor Llorà—, Tim Kovacs, Keiki Takadama, Pier Luca Lanzi, Wolfgang Stolzman, and Stewart Wilson edited a book *Advances at the frontiers of Learning Classifier Systems* which includes subject matter ranging from state-of-the art learning classifier systems (LCS) to the application of LCS on real-world problems from some of the leading researchers in the area.

Data-Intensive Technologies and Applications (DITA). Doctor Llorà co-founded with Doctor Michael Welge the Data-Intensive Technologies and Applications (DITA) research group at the National Center for Supercomputing Applications. The goal of the group is promote research and usage of techniques design to handle data-intensive computation. SAINT is a crucial part of DITA endeavors. More information can be found at <http://dita.ncsa.uiuc.edu/>.

10 New Discoveries, Inventions, or Patent Disclosures

Filled patent applications on new discoveries/inventions are under revision including:

- A robust and scalable GA, called design structure matrix GA, inspired from organizational theory.
- Collaborative systems design and innovation support over the web using interactive GAs, human-based GAs, and chance discovery.
- Evaluation relaxation for avoiding user fatigue and user inconsistency in interactive GAs.
- Efficient GA implementation for solving hard problems with millions and billions of variables
- Scalable methods for solving design, optimization, and modeling problems of complex systems with modularity, hierarchy, and overlap

In all, five patents have been filed in connection with AF research in the recent past:

Goldberg, D. E., T.-L. Yu, and A. Yassine. *Methods and Program Products for Optimizing Problem Clustering*. US utility patent pending (2004).

Goldberg, D. E., M. Welge, and X. Llorà, X. *Methods and Systems for Collaboration, Decision Support, and Knowledge Management*. US utility patent pending (2004).

Llorà, X., K. Sastry and D. E. Goldberg. *Methods and Systems for Interactive Computing*. US utility patent pending (2007).

Sastry, K., X. Llorà, and D. E. Goldberg. *A Method for Solving Large-Scale Optimization with More Than Millions of Variables Quickly, Reliably, and Accurately*. US provisional patent filing (2007).

Yu, T.-L., and D. E. Goldberg. *Using Genetic Algorithms, Dependency Structure Matrix, Substructural Chromosome Compression, and Proper Sequencing for Finding Extrema of Problems with Hierarchy and Overlap*. US provisional patent filing (2006).

11 Honors and Awards

JSD Professor. Professor Goldberg was named Jerry S. Dobrovolsky Distinguished Professor in Entrepreneurial Engineering in May 2003. The investiture was held September 23, 2003.

ISGEC Fellow. Professor Goldberg was named part of the inaugural class of Senior Fellows for the International Society for Genetic and Evolutionary Computation. (2003).

Other Awards. 1985 NSF Presidential Young Investigator. 1995, Associate, Center for Advanced Study (Illinois). 1996, Wickenden Award (ASEE). 1997, Gambrinus Fellow (Dortmund). NTU Outstanding Instructor Award (2001-2002).

Focused article gets picked by HPC wire. DISCUS's paper *Given Innovation a Boost* by Trish Barker was chosen by HPCwire editors to highlight new frontiers of computing. HPC wire, May 20, 2005.

HUMMIES-2007. During GECCO 2007 Dr. Llorà was awarded with two bronze awards for his work on bioinformatics.

Student Honors

GECCO-2006 best paper awards. Two papers from lab members won best paper awards at the 2006 Genetic and Evolutionary Computation conference.

GECCO-2007 best paper nominations. Four papers from lab members are nominated for best paper at the 2007 Genetic and Evolutionary Computation conference. The paper on the EDA track won the best paper award.

ICEIS 2007 best paper award. The paper presented by lab members at the ICEIS 2007 conference won the best paper award on Intelligent Systems and Distributed Agents.

GECCO 2008 best paper award nominees. Two papers by lab members have been nominated for best paper in the Estimation of Distribution Algorithms and Real World Applications tracks.